ESGF & NASA's Modeling Projects

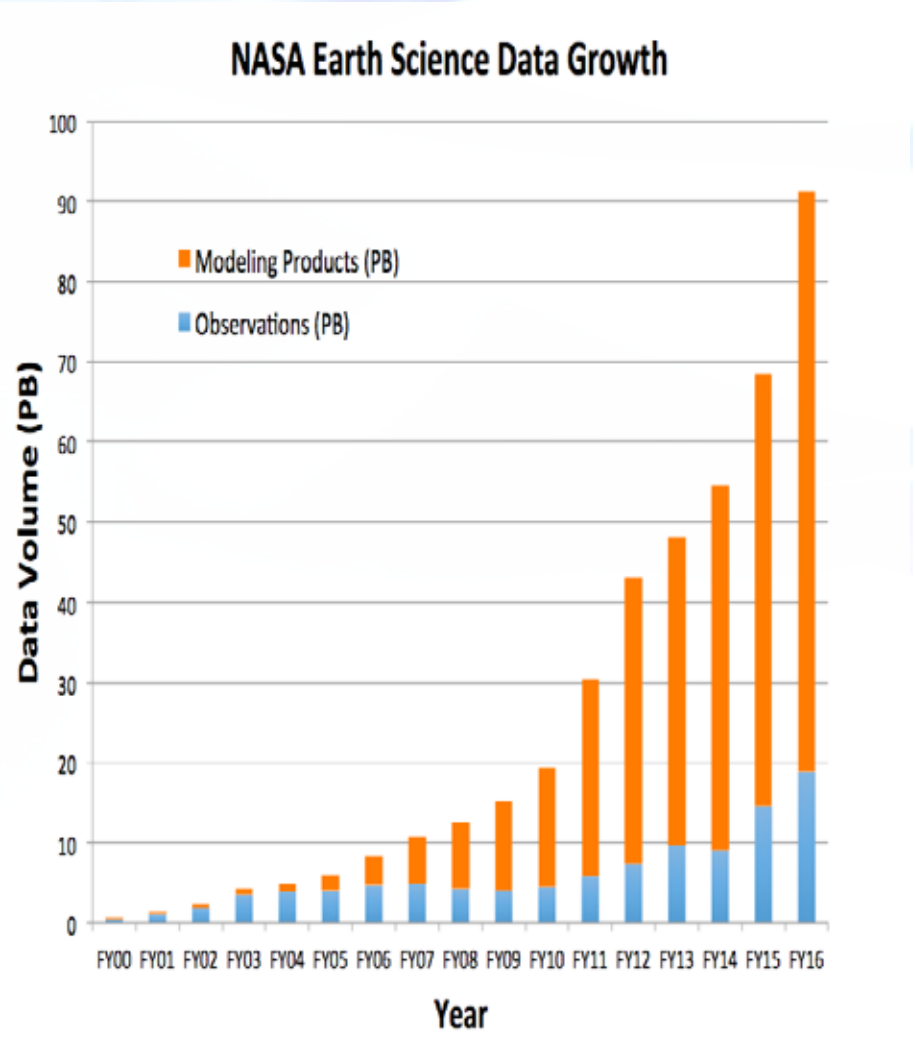
Tsengdar Lee

Presented at the 2016 ESGF F2F Meeting

December 6, 2016

Current NASA Earth Science Data Holding

- Observations include all products archived at EOSDIS (http://earthdata.nasa.gov/).
- Modeling products include all the high resolution climate modeling and data assimilation products at NASA Center for Climate Simulations (http://www.nccs.nasa.gov/) and NASA High-End Computing Capabilities (http://www.nas.nasa.gov/hecc/).
- Significant growth in modeling data is triggered by the availability of high resolution Earth observations and the computational resources.



2016 NASA Modeling Analysis & Simulation Product Plan

Forward Processing System	Satellite-Era Reanalysis 1979 - Present	EOS-Era Reanalysis 2000 – Present	Nature Runs (OSSEs)	Seasonal Forecast System	Coupled Simulations (Decadal, CMIP6)
3D-Hybrid Ensemble-Var (25km) 32 ensemble members Hydrostatic 1-Moment Cloud Microphysics *Current GEOS-5 FP system*	**MERRA (50km)** **Ending Feb. 2016** 3D-Var ~200 TB	**M2R12K (12km)** MERRA2 downscaled to 12 km Aerosols CO_2, CO, SO_2, O_3 Non-Hydrostatic 1-Moment Cloud Microphysics	**G5NR (7km)** Simulated 2005-2007 Aerosols, CO_2, CO, SO_2, O_3 Non-Hydrostatic 1-Moment Cloud Microphysics 4 PB	**GEOS SFS (50km)** MERRA-2 replay 50km, 40L ocean analysis 31 members per month Include aerosols, CO, CO_2 *M2-driven EnOI ocean analysis*	**GEOS CMIP (25km)** 25km Atmosphere 25km 50L ocean Include aerosols greenhouse gases Hydrostatic 2-Moment Cloud Microphysics *Planning/discussion and system evaluation in progress* *Will align with "MERRA-3" SFS and strategic direction o ESD*
3D-Hybrid Ensemble-Var (12km) 32 ensemble members Atmosphere, ocean surface Hydrostatic 2-Moment Cloud Microphysics *Parallel FP stream in 1Q-2016*	**MERRA-2 (50km)** 3D-Var Aerosols and CO, SO_2, O_3 1-Moment Cloud Microphysics ~400 TB	**IESA (12km)** 3D-Hybrid Ensemble-Var 32 ensemble members atmosphere, land, ocean surface Aerosols, CO_2, CO, SO_2, O_3 Non-Hydrostatic 2-Moment Cloud Microphysics 5,000 cores ; 40 simulation days/day 150 days total wallclock ~3 to 4 PB of data	**G5NR-CHEM (12km)** Simulated 2013-2014 Replay to M2R12K Full Reactive Chemistry Non-Hydrostatic 1-Moment Cloud Microphysics 1 PB of data *4Q-FY2016*	**GEOS SFS (25km)** Alignment with "MERRA-3" 25km, 50L ocean analysis System design under review *FY2019 target*	
4D Ensemble-Var (9km) ~100 ensemble members Atmosphere, ocean surface Non-Hydrostatic 2-Moment Cloud Microphysics (The first GEOS-6 system) *Parallel FP stream in 4Q-2016*	**MERRA-2 GMI replay (50km)** Replay GMI Chemistry 1 streams, 1,000 cores each 12 to 18 months ~ 1 PB	**IESAR4K (4km)** IESA Downscaled to 4km downscaling evaluation for NCA Aerosols, CO_2, CO, SO_2, O_3 Non-Hydrostatic 2-Moment Cloud Microphysics 5,000 cores ; 40 simulation days/day 150 days total wallclock ~3 to 4 PB of data	**G6NR (3km)** Simulated 2015 Aerosols CO_2, CO, SO_2, O_3, CH_4 Non-Hydrostatic 2-Moment Cloud Microphysics ~4 PB *Planning/evaluation*		
	Coupled Reanalysis ("MERRA-3") Atmosphere-land-ocean-cryosphere (alignment with SFS and CMIP6) *FY2019 target*				

Core GMAO projects completed, in-progress

Pathfinding projects toward GMAO core efforts.

FY16 Projects

Projects undergoing GMAO discussion/evaluati

Planned Future Projects

High Resolution Climate Projections

Climate Downscaling

DCP30 (Downscaled Climate Projections at 30arc sec)
Domain/Resolution: CONUS, ~800m
Frequency: Monthly
Variables: Tmax, Tmin, and Precip
No of CMIP5 models: 34
Baseline Data: Daly et al., 2002
Funding: NASA

BCCA (Bias Corrected Constructed Analogs)
Domain/Resolution: CONUS, ~12km
Frequency: Monthly
Variables: Tmax, Tmin, Precip
No of CMIP5 models: 21
Baseline Data: Maurer et al. 2002
Funding: USBR

LOCA (Localized constructed analogs)
Domain/Resolution: CONUS, ~6km
Frequency: Daily
Variables: Tmax, Tmin, Precip;
Humidity, Windspeed (in progress)
No of CMIP5 models: 32
Baseline Data: Livneh et al. 2013
Funding: USBR/CalEnergy

GDDP (Global Daily Downscaled Climate Projections)
Domain/Resolution: Global, ~25km
Frequency: Daily
Variables: Tmax, Tmin, and Precip
No of CMIP5 models: 21
Baseline Data: Sheffield et al. 2006
Funding: NASA

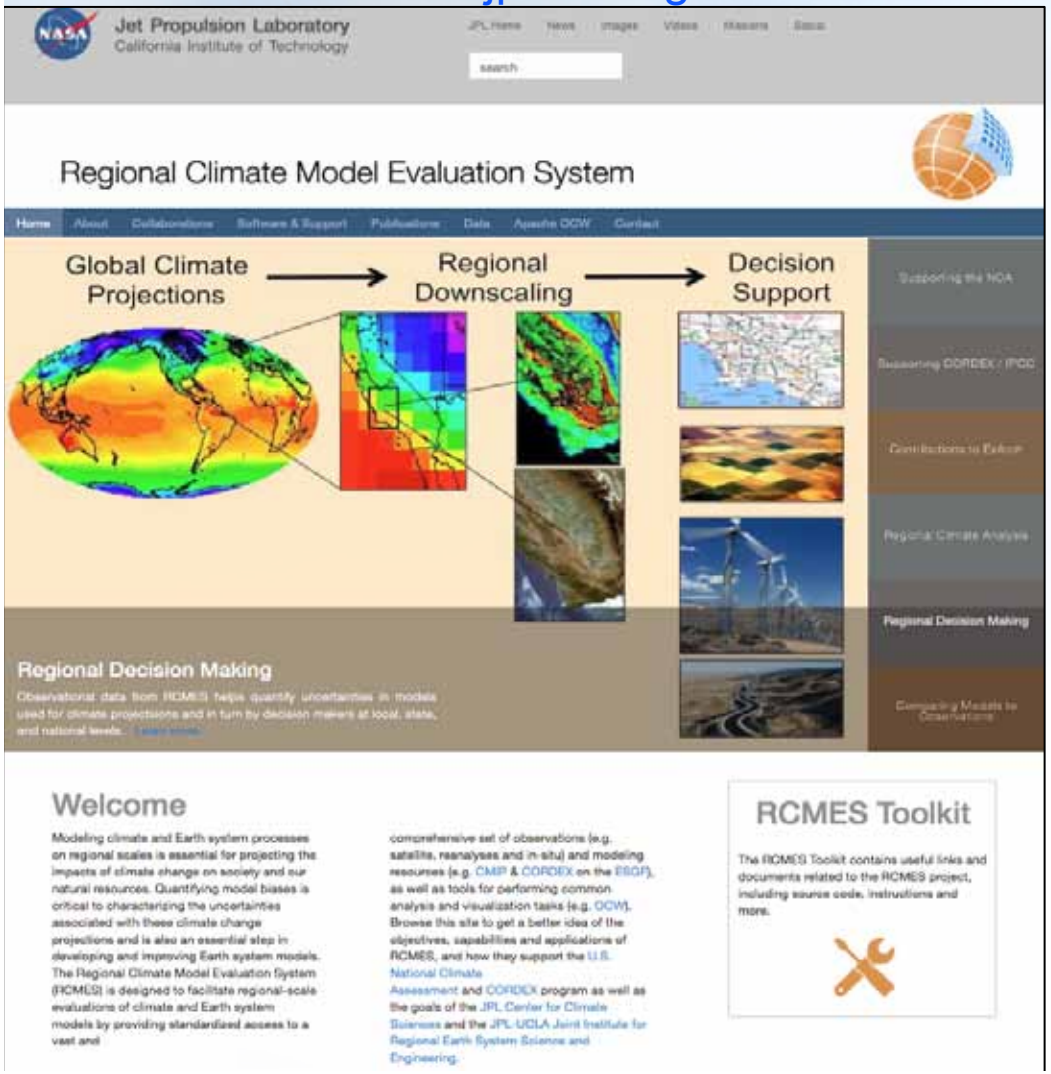
Gearing up for Climate Modeling Data Analytics

- Traditional data center focuses on data archive, access and distribution
  - Scientists typically order and download specific data sets to a local machine to perform analysis
  - With large amount of observational and modeling data, downloading to local machine is becoming inefficient
  - Data centers are starting to provide additional services for data analysis
- NASA computing and computational science program is building "data analytics platforms" using "Climate Analytics as a Service" (CAaaS) such as NASA Earth Exchange (**NEX**), Regional Climate Modeling Evaluation Systm (**RCMES**), Climate Model Diagnostic Analyzer (CMDA) and Observation for Model Intercomparison Project (Obs4MIPs) using Earth System Grid Federation (ESGF)
  - Build on technologies
  - Enabled by a rule based data management system
  - Current research focuses on how to manage data movement from the archives to the analytical platforms

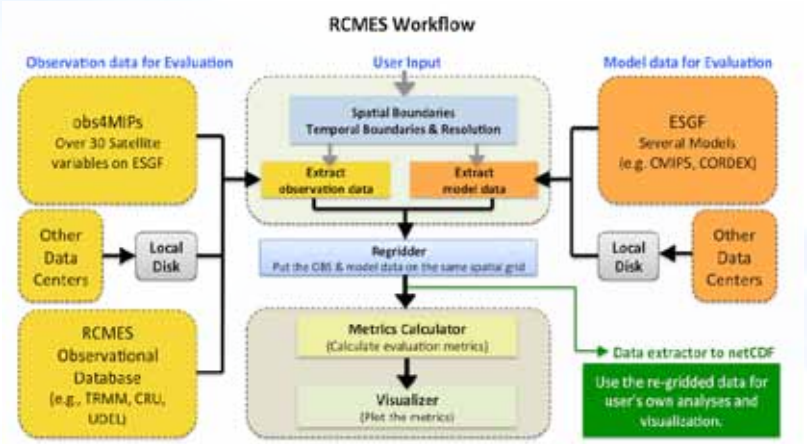
Regional Climate Model Evaluation System

RCMES facilitates regional model evaluation efforts via open source analysis toolkit and efficient links to model output (e.g. CORDEX, CMIP) and global observations (e.g. obs4MIPs, ana4MIPs).

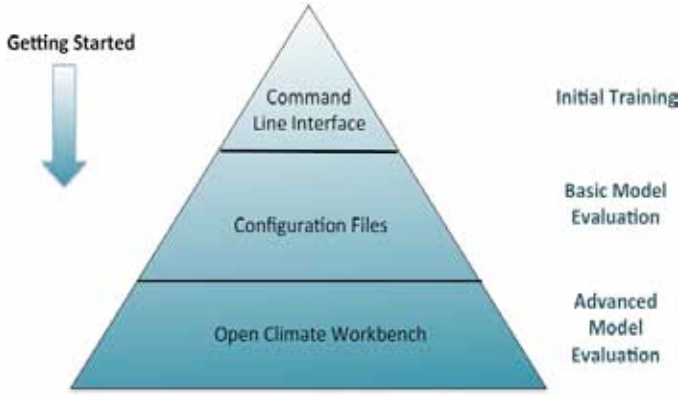
rcmes.jpl.nasa.gov



Modular and Open Source Design



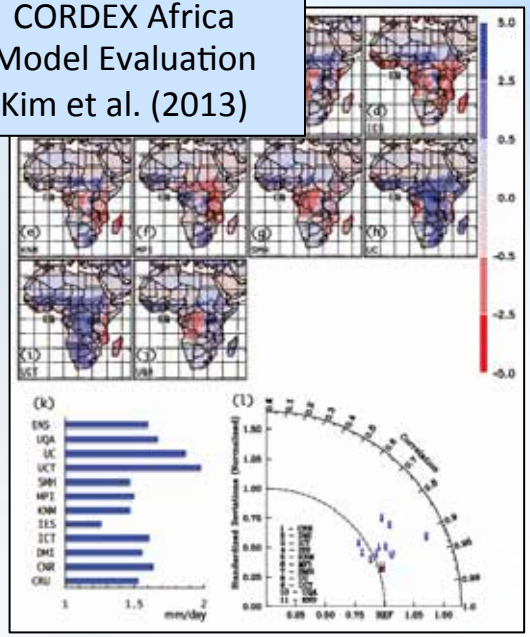
Graduated & Documented Training Materials



Regional Climate Model Evaluation System

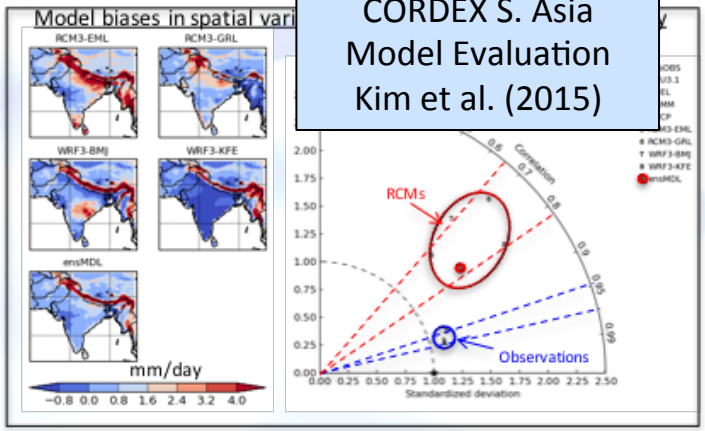
Contributing tools, resources and training to CORDEX

CORDEX Africa Model Evaluation Kim et al. (2013)



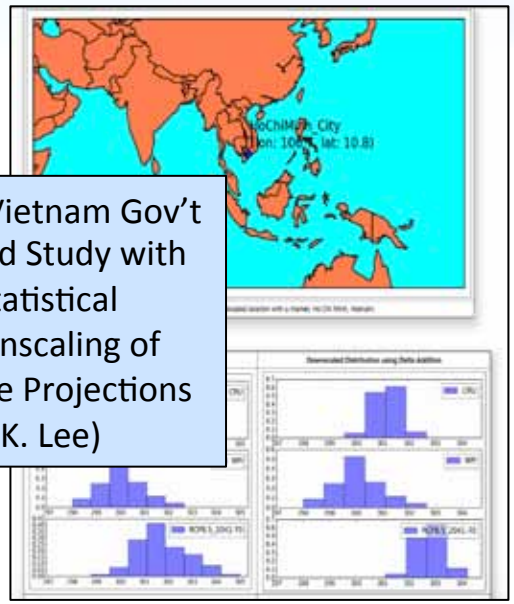
RCMES Article in WMO Bulletin (2012)

CORDEX S. Asia Model Evaluation Kim et al. (2015)



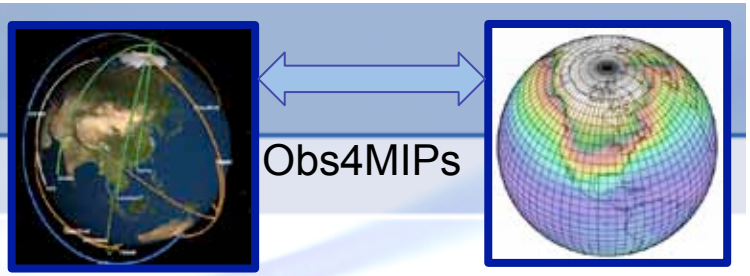
CORDEX-2016 Conference Stockholm, Sweden
RCMES Training; ~50 attendees

Assist Vietnam Gov't Funded Study with Statistical Downscaling of Climate Projections (K. Lee)

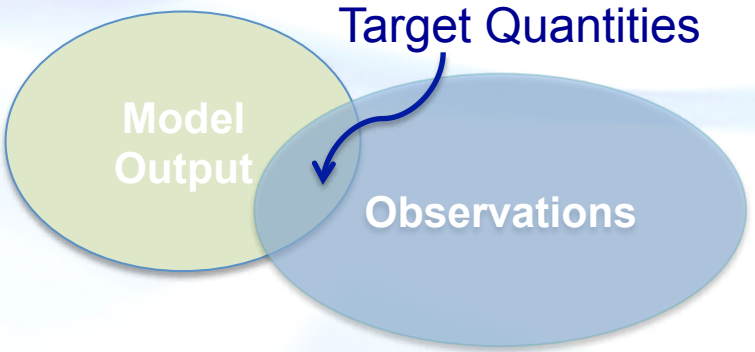


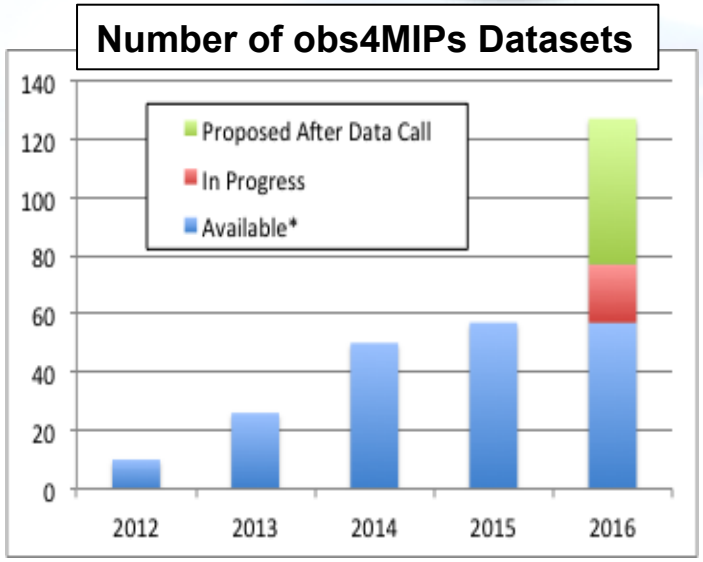
obs4MIPs

https://www.earthsystemcog.org/projects/obs4mips/



- A Project for identifying, documenting and disseminating observations for climate model evaluation.
- Data sets accessible on the Earth System Grid Federation (ESGF) alongside the Coupled Model Intercomparison Projection (CMIP) model output, adhering to the same data conventions, greatly facilitating research
- Guided by the World Climate Research Program (WCRP) Data Advisory Council (WDAC) obs4MIPS Task Team
- Growing international partnerships.



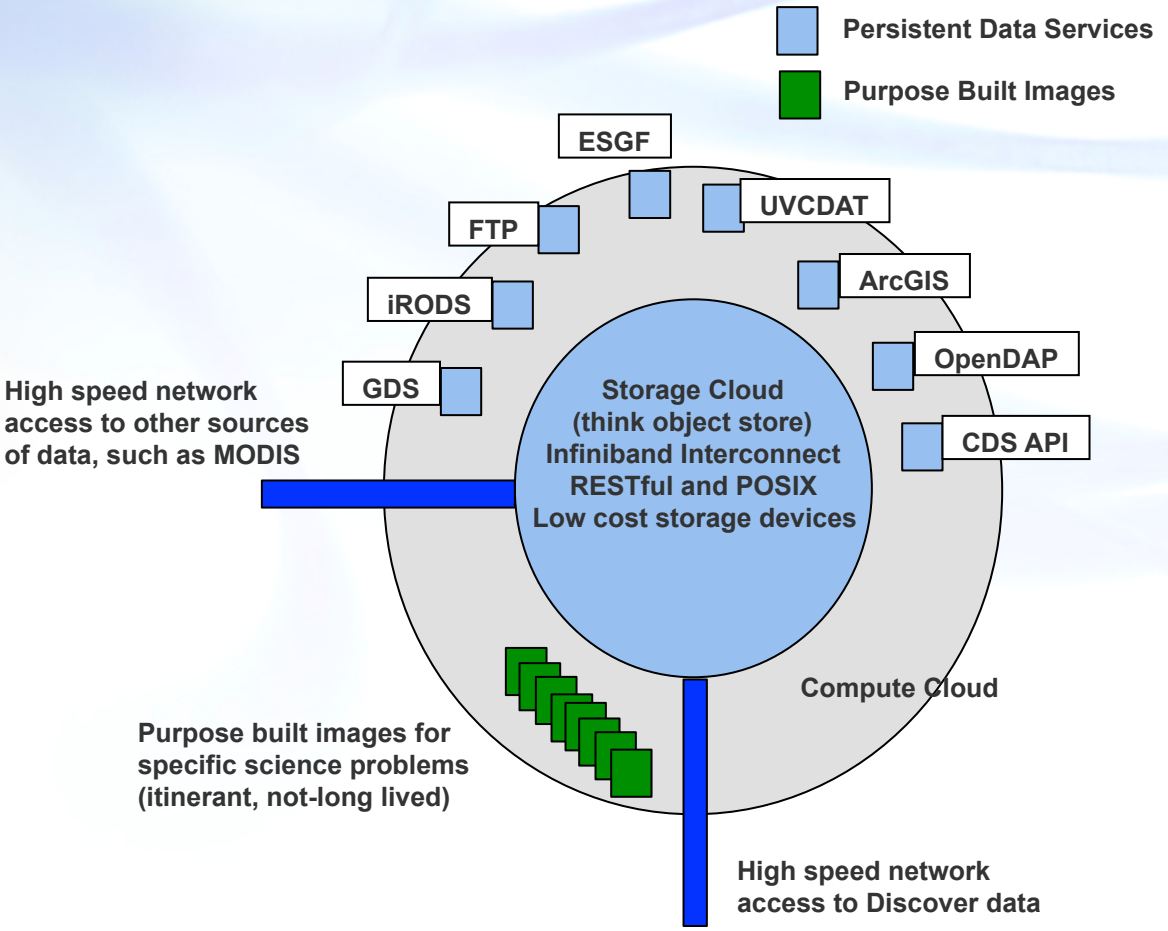


*ESGF is partially down until March

.... and growing!

Science Cloud Architecture

- Agile, high level of support
- Storage is 90% full prior to use
- The system owns the data
- The users own their analysis
- Extensible storage; build and expand as needed
- Persistent data services built in VMs, Containers, or bare metal
- Create purpose build VMs for specific science projects
- Image management



Climate Model Diagnostic Analyzer

- Web-based tools running on Amazon cloud.
- Only requirement from a user machine is a web browser with an internet connection. No local installation needed.
- Provides datasets and analysis services.
- You can analyze the datasets using the services.
- You can download analyzed output datasets.
- You can download original input datasets.

Major Challenges Over Next 10 Years and What Can We Do Now

- Challenge: Modeling and observational data will continue to grow exponentially
  - Major challenge in data management, analysis, and collaboration
  - Tape archives will not meet big data analysis challenges
  - Network will not catch up
  - Library model will no longer work
- Actions now:
  - Build centralized data analytics systems
  - Data proximal analytic capabilities (move the analytics to the data)
  - Commoditize data storage and data analytics
  - Explore and adopt new storage technologies (e.g., object storage)
- Large scale science informatics system will be needed to solve the future data challenges